Volume 16 Number 5 1988 Nucleic Acids Research

Second-strand cDNA synthesis with *E.coli* DNA polymerase I and RNase H: the fate of information at the mRNA 5′ terminus and the effect of *E.coli* DNA ligase

James M.D'Alessio and Gary F.Gerard

Molecular Biology Research and Development, Bethesda Research Laboratories, Life Technologies Inc., 8717 Grovemont Circle, Gaithersburg, MD 20877, USA

Received December 4, 1987; Accepted January 20, 1988

ABSTRACT

A simple method for generating cDNA libraries has been described (1) in which RNase H-DNA polymerase I-mediated second-strand cDNA synthesis primes from an RNA oligonucleotide derived from the 5' (capped) end of mRNA. The size of this oligonucleotide and the fate of the information corresponding to the RNA during subsequent cloning have not been established. We show here that the 5'-most RNA primer varies in length from 8 to 21 nucleotides, and that information corresponding to the length of the RNA primer is normally lost during cloning. A modification of the second-strand cDNA synthesis procedure is described which allows cloning of all, or almost all, of the primer sequence information. In addition, we show that the presence of *E. coli* DNA ligase during second-strand cDNA synthesis can increase the length of the cDNA clones obtained from long RNAs. Cloning by addition of linkers provides the greatest chance of obtaining near full-length cDNA clones from long mRNAs.

INTRODUCTION

The double-stranded (ds) cDNA synthesis method of Gubler and Hoffman (1) has become widely accepted. The method relies on the combined activities of *E. coli* RNase H and DNA polymerase I to catalyze second-strand cDNA synthesis from an mRNA-cDNA hybrid by RNA-primed nick translation (2). The method is attractive because it reduces the amount of hairpin double-stranded cDNA obtained by other methods (3), eliminating S1 nuclease-mediated cleavage of the hairpin loop. For the method to be successful, second-strand synthesis by DNA polymerase I must prime from the RNA oligonucleotide derived from the 5' (capped) end of the mRNA. Failure to prime from this oligonucleotide frequently leads to hairpin-primed synthesis of second-strand cDNA. However, once cDNA synthesis primes from the 5'-most RNA oligonucleotide, it is not clear whether sequence information within this RNA is preserved during subsequent cloning. We report here the details of the mechanism whereby the 5'-most RNA primer is removed during second-strand synthesis and the fate during subsequent cloning of the information derived from this RNA.

As originally conceived (2) and modified (1), DNA polymerase I-RNase H-mediated second-strand cDNA synthesis included E. coli DNA ligase to seal any nicks in the nick-translated second-strand cDNA. However, the need for DNA ligase has not been demonstrated experimentally (1). We report here the influence DNA ligase has on the size of second-strand cDNA products and on the size of cloned cDNA.

MATERIALS AND METHODS

RNAs

Synthetic 2.3-kb and 6.2-kb RNA containing a 19 nucleotide poly(A) tail at the 3'-end and an EcoR I cleavage site 38 nucleotides from the 5'-end were synthesized from linearized pHL1X and pHL3X, respectively (Figure 1). A typical reaction mixture (0.3 ml) contained 40 mM Tris-HCl (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine-HCl, 5 mM dithiothreitol (DTT), 0.4 mM each of ATP, CTP, UTP, and GTP, 20 μg/ml DNA, and 2000 U/ml T7 RNA polymerase. When 5'-end labeled RNA was synthesized, 0.4 mM [γ-^{32}P]GTP (20,000 cpm/pmole) was used. After a 1 hr incubation at 37°C, the RNA product was phenol extracted, ethanol precipitated, and purified by oligo(dT)-cellulose chromatography (3) to ensure the presence of a poly(A) tail.

First-strand cDNA synthesis

Reaction mixtures (50 μl) contained 50 mM Tris-HCl (pH 8.3 at 22°C), 75 mM KCl, 10 mM DTT, 3 mM $MgCl_2$, 500 μM each of dATP, dTTP, dCTP, and dGTP, 50 μg/ml oligo $(dT)_{12-18}$, 200 μg/ml RNA, and 10,000 units/ml cloned Moloney murine leukemia virus (M-MLV) reverse transcriptase (RT). When first-strand cDNA was labeled, [α-^{32}P]dCTP (250 cpm/pmole) was used. The reaction mixture was incubated at 37°C for 1 hr and chilled on ice. The yields were 40 to 60% and 30 to 40% from 2.3-kb and 6.2-kb RNA, respectively.

Second-strand cDNA synthesis

DNA polymerase I and RNase H combined. The first-strand reaction mixture was diluted to a final volume of 400 μl containing 25 mM Tris-HCl (pH 8.3), 100 mM KCl, 5 mM $MgCl_2$, 5 mM DTT, 250 μM each of dATP, dCTP, dGTP, and dTTP, 250 units/ml E. coli DNA polymerase I, and 8.5 units/ml E. coli RNase H. When second-strand cDNA was labeled, [α-^{32}P]dCTP (500 cpm/pmole) was used. Some reaction mixtures also contained 10 units/ml E. coli DNA ligase (New England Biolabs) and 0.15 mM β-NAD. After a 2 hr incubation at 16°C, the reaction mixture was extracted with phenol and the DNA was ethanol precipitated twice from ammonium acetate (2). The yields were generally 100%.

DNA polymerase I alone followed by RNase H. The first-strand reaction mixture was diluted and incubated under the reaction conditions described for second-strand cDNA synthesis, except that *E. coli* RNase H was omitted from the reaction. After phenol extraction and two ethanol precipitations, the double-stranded cDNA was incubated with *E. coli* RNase H for 20 min at 37°C in a reaction mixture (20 μl) containing 20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 20 mM KCl, 0.1 mM Na_2 EDTA, 0.1 mM DTT, and 20 units/ml *E. coli* RNase H. The DNA was phenol extracted and ethanol precipitated.

Cloning

Cloning by linker addition. To ensure the ends of the double-stranded cDNA were blunt, the DNA was incubated with T_4 DNA polymerase and all four dNTPs (4). *Hin*d III linkers were ligated to blunt-ended double-stranded cDNA (5). Gel filtration over a Sepharose 4B-Cl column (0.7 x 4 cm and 0.2 x 30 cm for cDNA synthesized from 2.3-kb and 6.2 kb-RNA, respectively) was used to eliminate excess linkers and low molecular weight cDNA prior to ligation to plasmid. The cDNA was ligated to pUC8 or pUC9 DNA that had been cut with *Hin*d III and dephosphorylated with bacterial alkaline phosphatase using standard techniques (5). The ligation mixture was diluted and used to transform *E. coli* DH5α competent cells (BRL) to ampicillin resistance. The cells were plated onto plates containing 5-bromo-4-chloro-3-indoly-β-D-galactopyranoside (X-gal). White colonies were picked and grown overnight. The DNA was extracted (6), digested separately with *Eco*R I and *Hin*d III, and fractionated by electrophoresis in a 1.2% agarose gel to determine the size of the cDNA insert.

Cloning by tailing with terminal deoxynucleotidyl transferase. Double-stranded cDNA synthesized from 6.2-kb RNA, but not 2.3-kb RNA, was size fractionated on a Sepharose 4B-Cl column (0.2 x 30 cm) to eliminate product less than 500 base pairs in length prior to tailing. DNA was tailed with dCTP (7,8) in reaction mixtures (20 μl) containing 100 mM potassium cacodylate (pH 7.2), 2 mM $CoCl_2$, 0.2 mM DTT, 20 μM [^{3}H]dCTP (2000 cpm/pmole), cDNA equivalent to 1 pmole 3' termini, and 1000 units/ml terminal deoxynucleotidyl transferase (TdT). The reactions were incubated for 15 min at 30°C, and the average dC-tail length ranged from 20 to 30 nucleotides.

The dC-tailed ds cDNA was annealed to *Pst* I - cut pBR322 DNA in 100 μl containing 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.1 mM Na_2 EDTA, 25 ng dG-tailed pBR322 DNA, and either 8 to 30 ng 2.3-kb cDNA or 21 to 82 ng 6.2-kb cDNA. Mixtures were incubated at 68°C for 5 min and 45°C for 60 min. A

portion of the annealing mixture was used to transform E. coli DH5α competent cells (BRL) to tetracycline resistance. Resistant colonies were grown overnight in minimal media containing 15 μg/ml tetracycline. The DNA was extracted (6), digested with EcoR I and Pst I alone or in combination, and fractionated by electrophoresis in a 1.2% agarose gel to determine the size of the cDNA insert.

5'-End analysis of second-strand cDNA

Unlabeled, apparent full-length, double-stranded cDNA synthesized from 2.3-kb RNA in the absence of E. coli RNase H was purified by agarose gel electrophoresis (9), dephosphorylated with bacterial alkaline phosphatase (10), and labeled at 5' ends with [γ-^{32}P]ATP and T_4 polynucleotide kinase (11). After digestion with EcoR I and repair with dTTP, dATP, dCTP, dGTP and DNA Polymerase I Large Fragment, labeled 40-43 nucleotide fragments containing the 5'-most RNA primer covalently joined to second-strand cDNA were separated by electrophoresis on a 10% polyacrylamide gel (11). The identity of the labeled nucleoside 5'-monophosphate at the 5' end of each fragment was established by digestion with snake venom phosphodiesterase (12) (Worthington) followed by thin-layer chromatography on PEI-cellulose in 1 M LiCl (13) to separate nucleotides.

RESULTS AND DISCUSSION

RNAs used as templates

Synthetic 2.3-kb and 6.2-kb RNAs with a 3'-poly(A) tail, synthesized from plasmid pHL1X and pHL3X, respectively (Figure 1), were used as model templates. These RNAs were selected because they can be synthesized in vitro in abundance; they can be labeled uniformly or at an end during synthesis; the sizes of the RNAs, 2.3 kb and 6.2 kb, are typical of an average size and a long mRNA, respectively; and they contain a unique EcoR I cleavage site 38 bases from the 5' end, which can be exploited to establish the details of the second-strand cDNA priming event and to assess the length of cloned cDNA inserts (see below).

Size of the 5'-most RNA oligonucleotide generated during first-strand cDNA synthesis

Complete retention of sequence information during double-stranded cDNA synthesis using Gubler and Hoffman (1) methodology requires synthesis of full-length first-strand cDNA and priming of second-strand synthesis from an RNA oligonucleotide derived from the 5' end of the mRNA. This 5'-most RNA oligonucleotide could be generated from the mRNA-cDNA hybrid either by the

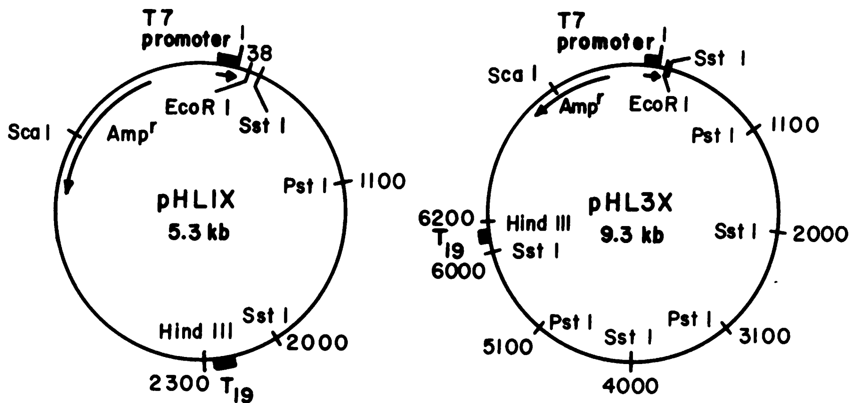

Figure 1. The structure of plasmids pHL1X and pHL3X used to synthesize 2.3-kb and 6.2-kb RNA, respectively. Plasmid DNA was digested with Hind III before transcription with T7 RNA polymerase (Materials and Methods).

action of RT-associated RNase H after completion of first-strand synthesis or by the action of E. coli RNase H added at the beginning of the second-strand reaction. At the end of first-strand cDNA synthesis catalyzed by M-MLV RT, much of a 2.3-kb 5'-end labeled RNA template has been degraded, and the 5'-most oligonucleotides generated are approximately 10 to 20 nucleotides long (Figure 2). These oligonucleotides are long enough to remain hydrogen bonded to first-strand cDNA and would prime second-strand synthesis without further cleavage by E. coli RNase H. In addition, with RNA oligonucleotides of this length, the 5'-most priming event of second-strand cDNA synthesis would occur upstream from the EcoR I site in the 2.3-kb RNA sequence, ensuring that this site would be available for cleavage in full-length ds cDNA product (see below). Similar size fragments were generated during first-strand synthesis catalyzed by AMV RT (data not shown).

Fate of the 5'-most mRNA sequence information

Experimental approaches. To study the fate of the 5'-most RNA oligonucleotides that prime second-strand synthesis, ds cDNA product from 2.3-kb RNA was cut with EcoR I, and the 3'-recessed ends at the EcoR I site were labeled by repair synthesis with DNA polymerase I Large Fragment (Figure 3). Because of the location of the EcoR I site (Figure 1), only near full-length ds cDNA molecules were cut. Repair synthesis at the EcoR I termini labeled (a) the 3' ends of short fragments of second-strand cDNA (less than 45 nucleotides long) that contained at their 5' ends any

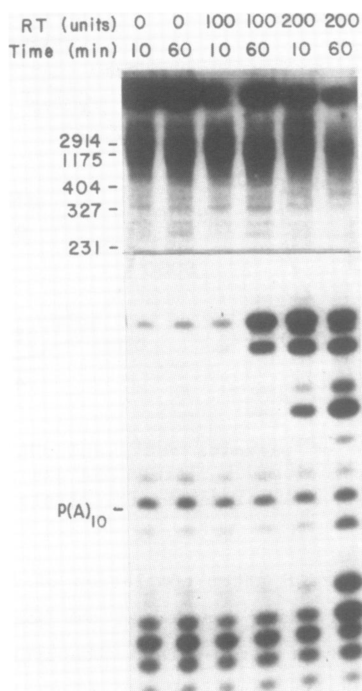

Figure 2. Electrophoretic analysis of ^{32}P-labeled RNA oligonucleotides generated during cDNA synthesis from 5'-end labeled 2.3 kb RNA. First-strand cDNA synthesis reactions (25 μl) containing 1 μg (18,000 cpm) of 2.3 kb RNA labeled with [γ-^{32}P]GTP, the indicated amounts of M-MLV RT, and [^{3}H]dGTP (70 cpm/pmole) to monitor DNA synthesis, were incubated at 37°C. Duplicate aliquots (2.5 μl) were removed at the times indicated. One was electrophoresed on a 7 M urea - 10% polyacrylamide gel and the other was acid precipitated to determined amounts of cDNA synthesized (0.28 and 0.93 μg cDNA were synthesized in 60 min with 100 and 200 units of M-MLV RT, respectively). Virtually no RNA migrated in the 25 to 230 nucleotide size range. This section of the gel is not shown.

surviving RNA primer and (b) the 3' end of a long fragment of first-strand cDNA (~ 2,200 nucleotides long). Repair synthesis also labeled other regions in ds cDNA products, e.g., gaps or 3'-recessed ends derived from the poly(A) end of the mRNA. DNAs labeled at these locations were usually greater than 100 nucleotides long and did not interfere with the analysis (Figure 4, lane H). The shorter fragments labeled at the EcoR I site and generated under several different second-strand reaction conditions (Figure 3) were size fractionated and separated from other labeled fragments by electrophoresis on a denaturing sequencing gel (11) (Figure 4). The structure of these shorter fragments will be discussed in the next sections.

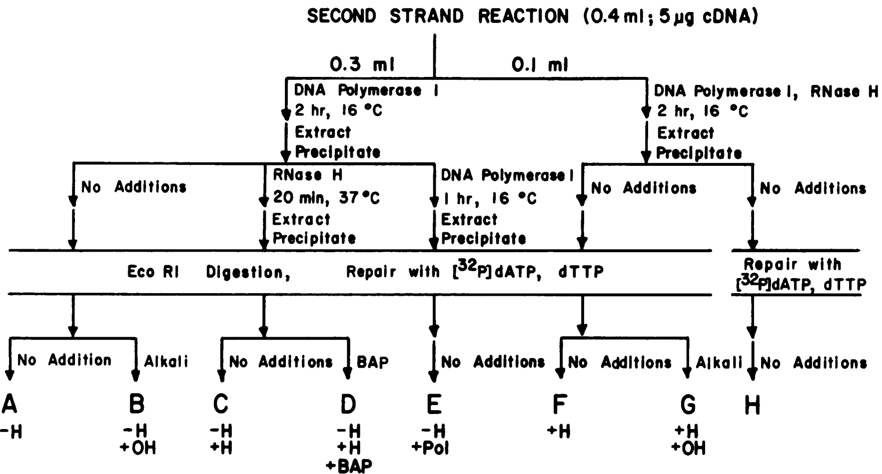

Figure 3. Schematic of protocol used to prepare double-stranded cDNA synthesized from 2.3 kb RNA for electrophoretic analysis (Figure 4). First-strand cDNA synthesis was performed as described in Materials and Methods from 2.3 kb RNA. Yields were 50% and half the product was apparent full-length cDNA. On ice, the reaction mixture was diluted to 400 µl in second-strand components (Materials and Methods), which included DNA polymerase I but excluded RNase H and DNA ligase, was split into two aliquots (0.3 ml and 0.1 ml), and E. coli RNase H (0.5 units) was added to the 0.1 ml aliquot. Both reaction mixtures were incubated at 16°C for 2 hr, phenol extracted, and ethanol precipitated. Products were then subdivided into portions and treated as indicated. RNase H digestion (indicated by +H) was performed with 0.5 units at 37°C for 20 min (Materials and Methods). Incubation with DNA polymerase I (indicated by +Pol) was for 1 hr at 16°C at 250 units/ml (Materials and Methods). All cDNA products were labeled by DNA repair synthesis with [α-^{32}P]dATP, dTTP, dCTP, dATP, and DNA polymerase I Large Fragment after digestion with EcoR I (the DNA in H was labeled without digestion). The completeness of the repair reaction was confirmed by cutting the labeled DNA at the Cla I site located just 5′ to the EcoR I site and showing that the labeled DNA fragment migrated during gel electrophoresis as a single species 11 nucleotides long. Treatment with alkali (indicated by +OH) was in 0.3 N KOH at 37°C for 16 hr. Treatment with bacterial alkaline phosphate (indicated by +BAP) was done at 20,000 units/ml for 30 min at 65°C.

In a typical experiment, after EcoR I digestion and repair, approximately 25% of the label resided in the shorter second-strand cDNAs, 25% resided in the 2.2 kb first-strand cDNA, and 50% was found in DNA labeled at locations other than at the EcoR I site.

To study the fate of the sequence information corresponding to the 5′ end of the mRNA, double-stranded cDNA products were cloned and the size of the cloned cDNA inserts were determined. In some cases, the DNA sequence upstream from the EcoR I site of the cloned insert was determined.

A

M A B C D E F G H

-H; -H +OH; -H +H; -H +H +BAP; -H +Pol; +H; +H +OH

89- 56- 35- 22-

Primer + cDNA 43- 40

—65 Hairpin

28- 25- 24- 23-

-27 -26 -25 -27° -26° -25°

-35° -34° -33° -32° -30° -29°

B

A and E	*(G)(G)(G)AGACCACAACGGUUUCC*CTCTAGCGGGATCGATCCGA*A*TT	40-43 nt
B	CTCTAGCGGGATCGATCCGA*A*TT	23 nt
C	p*CC*CTCTAGCGGGATCGATCCGA*A*TT	25 nt
D	*CC*CTCTAGCGGGATCGATCCGA*A*TT	25 nt
F and G	pCACAACGGTTTCCCTCTAGCGGGATCGATCCGA*A*TT	35 nt

C

```
 T   C
T G GGT TCCCtctagcgggatcgatccgaatt  3'
 T  CCA AGGGAGATCGCCCTAGCTAGGCTTAA  5'   65 nt
  G    A
```

Second-strand synthesis in the absence of RNase H. If second-strand cDNA was synthesized by DNA polymerase I in the absence of E. coli RNase H, a family of labeled fragments 40 to 43 nucleotides long was observed (Figure 3A and Figure 4, section A, lane A). Treatment of these fragments with alkali increased their rate of migration (Figure 3B and Figure 4, section A, lane B), demonstrating the presence of RNA in the fragments. The structures of the major oligonucleotide product present before and after treatment with alkali are shown in Figure 4, section B, line A and B, respectively. The DNA that remained varied in length from 23 to 28 nucleotides. Therefore, second-strand cDNA synthesis in the absence of E. coli RNase H primed primarily 23 to 28 nucleotides upstream from the EcoR I site corresponding to 21, 20, 19, and 16 nucleotides from the original 5′ end of the 2.3-kb RNA. These deduced primer lengths are consistent with the length of the 5′-end labeled RNA oligonucleotides generated by the RNase H activity of M-MLV RT and found at the end of the first-strand reaction (Figure 2).

The length heterogeneity of these fragments before alkali treatment (40 to 43 nucleotides) was due to the absence of one, two, or three nucleotides from the 5′ end of the fragments. This was established by identifying the 5′-end nucleotide of each fragment separated by electrophoresis on a 10% polyacrylamide gel (Materials and Methods).

Figure 4. (A) Electrophoretic analysis of double-stranded cDNA. Double-stranded cDNAs prepared and labeled as described in Figure 3 were adjusted to 10,000 cpm/μl in 8 M urea and were denatured by heating at 90°C for 30 sec. Equal volumes of DNAs were fractionated on an 8.3 M urea, 10% polyacrylamide sequencing gel to determine the size of the radiolabeled fragments derived from the 5′ end of second-strand cDNA. Lanes A-H correspond to samples A-H in Figure 3 (H: E. coli RNase H; OH: alkali; BAP: bacterial alkaline phosphatase; Pol: E. coli DNA polymerase I). The numbers indicate the length of the fragments in nucleotides. Those numbers followed by a (O) indicate that the fragments contain 5′ phosphate and cannot be sized directly with the markers used. Lane M is labeled, dephosphorylated, Taq I-cut ϕX174 RF DNA. DNA labeled by repair of the other half of the EcoR I cut site remained at the top of the gel and is not shown. (B) Structure of the most abundant labeled fragments observed in lanes A-G. Ribonucleotides are in italics and asterisks designate radiolabeled deoxyribonucleotides. (C) Sequence and proposed structure of the 65-nucleotide hairpin cDNA observed towards the top of the gel lanes. The hairpin cDNA was prepared as follows. Unlabeled first-strand cDNA was treated with alkali to remove RNA, incubated with DNA polymerase I in a second-strand reaction, digested with EcoR I, labeled by repair synthesis as described, and fractionated by gel electrophoresis. The labeled band that comigrated with authentic 65-nucleotide hairpin cDNA synthesized in a standard second-strand reaction (Materials and Methods) was excised and eluted from the gel and sequenced (11). Putative first-strand cDNA is represented in capital letters and second-strand cDNA in lower case letters.

Table 1. Cloning Efficiencies

Vector + Double-Stranded cDNA Insert	Efficiencies Recombinants / µg cDNA Insert	Recombinants[f] / µg mRNA
dG-Tailed pBR322 DNA + dC-Tailed 2.3 kb cDNA[a,b]	0.8×10^6	0.4×10^6
dG-Tailed pBR322 DNA + dC-Tailed 2.3 kb cDNA[a,c]	5.3×10^6	2.3×10^6
pUC8 DNA + 2.3 kb cDNA with Hind III Linkers[d,e]	2.2×10^6	0.3×10^6

[a]Supercoiled pBR322 DNA gave a transformation efficiency of 6.7×10^7 colonies/µg DNA, while dG-tailed pBR322 DNA gave 1.4×10^4 colonies/µg DNA
[b]Double-stranded cDNA was synthesized in the absence of *E. coli* RNase H
[c]Double-stranded cDNA synthesized in the absence of *E. coli* RNase H was incubated with RNase H before tailing
[d]Supercoiled pUC DNA gave a transformation efficiency of 2×10^8 colonies/µg DNA. Background (blue colonies) was 7% during transformation of the ligated mixture
[e]Double-stranded cDNA was synthesized in the presence of *E. coli* RNase H
[f]Recombinants/µg mRNA = Recombinants/µg cDNA insert X $\frac{\text{\% cDNA yield}}{0.5(100)}$

It cannot be established from the data presented here when ribonucleotides were lost from the 5′ end of the RNA primer. The 5′ → 3′ exonuclease activity of DNA polymerase I is capable of hydrolyzing RNA

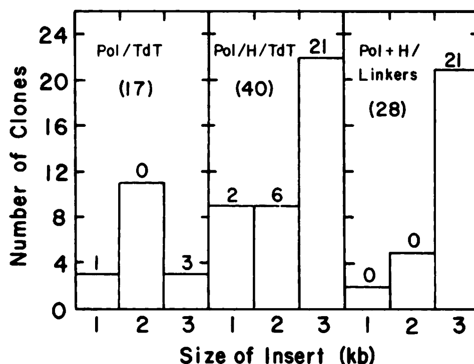

Figure 5. Size of the cDNA inserts synthesized from 2.3 kb RNA and cloned by different methods. First-strand cDNA synthesized from 2.3 kb RNA was made double-stranded with DNA polymerase I in the absence of *E. coli* RNase H (left), by subsequent incubation with RNase H (middle), or by simultaneous incubation with RNase H (right) (Materials and Methods). DNA was cloned into pBR322 DNA by tailing with TdT or into pUC8 by addition of *Hind* III linkers (Materials and Methods). The sizes of the cDNA inserts (1 = 0 to 1 kb; 2 = 1 to 2 kb; 3 = 2 to 3 kb) were established as described in Materials and Methods. The numbers in parentheses are the total number of clones screened. The numbers above the bars indicated the number of cDNA inserts in each size class that contained an *EcoR* I cut site.

contained in an RNA-DNA hybrid (14). However, DNA polymerase I did not catalyze the removal of these nucleotides. As shown in Figure 4 (compare lanes A and E), incubation of the ds cDNA product synthesized in the absence of RNase H with additional DNA polymerase I caused no change in the size distribution of this family of fragments. As will be shown below, the RNA primer present in a hybrid with first-strand cDNA is attacked by E. coli RNase H, and once degradation is started by this enzyme, the DNA polymerase I 5′ → 3′ exonuclease can complete removal of the RNA primer.

RNA at the 5′ end of first-strand DNA in this product precluded cloning by ligation of DNA linkers. As an alternative, the DNA was cloned by adding dC homopolymer tails with TdT and annealing the tailed DNA to dG-tailed pBR322 DNA. The cloning efficiency was relatively low (4×10^5 recombinants/μg in mRNA, Table 1), and only a small proportion (3 of 17) of the cloned inserts examined was near full-length (contained an EcoR I site) (Figure 5). This means that the 3′ end of full-length or near full-length first-strand DNA in this cDNA product was not tailed efficiently by TdT, probably because it was recessed and/or was part of a blunt ended DNA-RNA hybrid (7,8).

Second-strand synthesis with a delayed addition of RNase H. If E. coli RNase H was added to the second-strand reaction after DNA polymerase I was removed, some, but not all, of the RNA primer was removed (Figure 3C and Figure 4, section A, lane C). Since E. coli RNase H leaves a 5′ phosphate after cleavage (15), the 5′ phosphate was removed to allow direct size comparison with the alkali treated cDNA fragments (Figure 3D and Figure 4, section A, lane D). After phosphatase treatment, the mobility of each band decreased by an amount corresponding to approximately 1½ nucleotides, and each band migrated at a position consistent with a length 2 nucleotides greater than when all RNA had been removed with alkali (Figure 4, compare lanes B and D). Therefore, E. coli RNase H removed all but two ribonucleotides of the RNA primer. This corroborates the conclusions of others (15-17) that E. coli RNase H is not capable of cleaving hybrid RNA at the junction between a ribonucleotide and a deoxyribonucleotide. The structures of the major oligonucleotide product present before and after treatment with bacterial alkaline phosphatase are shown in Figure 4, section B, line C and D, respectively.

Removal of most of the RNA primer of second-strand cDNA with E. coli RNase H should have made the 3′ end of full-length first-strand cDNA single-stranded, and therefore, a good substrate for tailing with TdT. After

addition of dC homopolymer tails, the cDNA was annealed to dG-tailed pBR322 DNA and transformed into E. coli. The cloning efficiency increased 6.5 fold relative to cDNA not treated with RNase H (Table 1), and a large proportion (21 of 40) of the cloned inserts was near full-length (Figure 5). Two apparent full-length 2.3-kb cDNA inserts were sequenced from the EcoR I site through the dC-dG homopolymer region (11). The clones had the complete sequence, and differed only in the length of the tail. Because the 5′ end of the RNA starts with 3 successive G residues (see Figure 4, section B), it was impossible to determine which complementary C residue at the 3′ end of first-strand cDNA served as the initiation point for dC-homopolymer tail synthesis. It is clear that no more than 3 nucleotides corresponding to the 5′ end of the RNA were lost during cDNA synthesis and cloning by this method.

Second-strand synthesis in the presence of RNase H. If E. coli RNase H was present during DNA polymerase I-catalyzed second-strand cDNA synthesis, a family of longer fragments with a broad size distribution was observed (Figure 3F and Figure 4, section A, lane F). Since treatment with alkali did not change the migration rate of most of these fragments (Figure 3G and Figure 4, section A, lane G), all the RNA primer had been removed during second-strand synthesis. We conclude that RNase H initiated degradation of the primer, followed by 5′ → 3′exonucleolytic digestion by DNA polymerase I, removed all the ribonucleotides. The fragments were accurately sized after treatment with bacterial alkaline phosphatase, and the majority were 29 to 35 nucleotides long (data not shown). Therefore, second-strand cDNA synthesis in the presence of E. coli RNase H primed primarily 14, 13, 11, 10, 9, and 8 nucleotides from the original 5′ end of the 2.3-kb RNA. Apparently, the 5′-most RNA oligonucleotides (16 to 21 nucleotides long) generated by M-MLV RT RNase H during first-strand cDNA synthesis were shortened at their 3′ ends by E. coli RNase H before second-strand synthesis was initiated.

Similar gel analyses were performed with rabbit α-globin double-stranded cDNA cut with Sty I and repaired (data not shown). When E. coli RNase H was absent during second-strand synthesis, the capped RNA primer was predominately 12 to 18 nucleotides long; in the presence of E. coli RNase H, the RNA primer length was 7 to 13 nucleotides.

Removal of the 5′-most RNA primer by the combined action of E. coli RNase H and DNA polymerase I would make any complementary nucleotides located at the 3′ end of full-length first-strand cDNA single-stranded.

Single-stranded DNA in turn would be hydrolyzed by the 3' → 5' exonuclease activity of DNA polymerase I, leaving the end of the double-stranded cDNA product blunt, or nearly so. Therefore, during subsequent cloning, sequence information corresponding to the 5'-most RNA oligonucleotide that primed second-strand synthesis would be lost. To test this, Hind III linkers were ligated to ds cDNA synthesized from 2.3-kb RNA in the presence of *E. coli* RNase H and the cDNA was cloned into the Hind III site of pUC8. Because of the low overall yield of cDNA obtained using this method, the cloning efficiency was only 3 x 10^5 recombinants/μg mRNA (Table 1). However, a high proportion of cDNA inserts examined were near full-length (21 of 28) (Figure 5). Two apparent full-length inserts were sequenced from the EcoR I site through the Hind III linker and into the vector (11). One was missing 12 nucleotides, and the other 15 nucleotides, corresponding to the 5' end of the RNA sequence. These results suggest that the sequence information lost following cloning corresponds to the sequence of the 5'-most RNA oligonucleotide that primed second-strand synthesis.

Hairpin-primed second-strand synthesis. Second-strand synthesis of cDNA by all three methods just described is accompanied by a small amount of hairpin-primed synthesis. Usually, 5 to 15% of second-strand cDNA is synthesized from a hairpin as judged by electrophoretic analysis of the cDNA (data not shown). The fragment 65 nucleotides long observed in lanes A through G in Figure 4 could be generated by hairpin synthesis, since no RNA was found associated with it, even in the absence of *E. coli* RNase H. This DNA fragment was synthesized in large quantities, labeled, isolated, and sequenced (Figure 4, section C). The sequence suggests that second-strand synthesis primed from the 3' end of full-length first-strand cDNA that had annealed to a complementary stretch of 7 nucleotides with one mismatch (Figure 4, section C). S1 nuclease treatment of this DNA and subsequent cloning by linker addition would probably result in the loss of 17 nucleotides corresponding to the 5' end of the RNA.

Effect of DNA ligase on second-strand cDNA synthesis

No systematic study has been carried out demonstrating the influence of *E. coli* DNA ligase on either the size of second-stand cDNA synthesized by DNA polymerase I and RNase H from RNA templates of different sizes or the size of cDNAs which are recovered in clones. The proportion of full-length second-strand cDNA synthesized from RNA decreases as the length of the RNA increases. For example, 50 to 60% of the second-strand cDNA synthesized from 2.3-kb RNA was full-length, as judged by agarose gel electrophoresis,

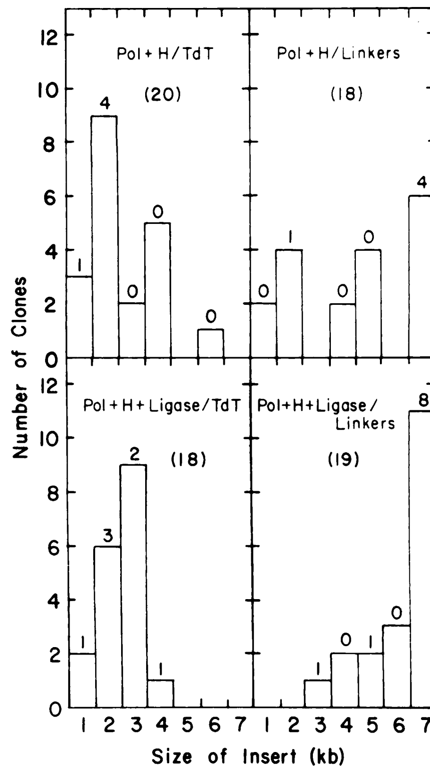

Figure 6. Size of the cDNA inserts synthesized from 6.2 kb RNA and cloned by different methods. First-strand cDNA synthesized from 6.2 kb RNA was made double-stranded with DNA polymerase I and RNase H in the absence (upper panels) and presence (lower panels) of E. coli DNA ligase. DNA was cloned into pBR322 DNA by tailing with TdT (left panels) or into pUC9 by addition of Hind III linkers (right panels). The sizes of the cDNA inserts were established as described in Materials and Methods. The numbers in parentheses are the total number of clones screened. The numbers above the bars indicate the number of cDNA inserts in each size class that contained an EcoR I cut site.

while only 10% synthesized from 6.2-kb RNA was full-length (data not shown). In both cases, the presence of E. coli DNA ligase did not increase this percentage.

In the original vector-primed cDNA synthesis procedure of Okayama and Berg (2), E. coli DNA ligase was required during second-strand synthesis to successfully clone full-length globin cDNA. However, if first-strand cDNA synthesis from globin mRNA was primed with oligo(dT) rather than dT-tailed plasmid, E. coli DNA ligase was not required during second-strand synthesis

to clone near full-length globin cDNA (1). When second-strand cDNA was synthesized from 6.2-kb RNA with DNA polymerase I and RNase H, tailed with dCTP and TdT, and cloned in dG-tailed pBR322 DNA, most cloned inserts were less than 4 kb long (Figure 6). Addition of E. coli DNA ligase to the second-strand reaction did not substantially increase the length of the cloned inserts. Interestingly, many of the clones (25 to 39%) generated by tailing with TdT contained inserts with an EcoR I restriction site. These clones were probably generated from a cDNA product in which the first-strand cDNA was near full-length (synthesis had proceeded to at least within 38 bases of the 5′ end of the RNA) and the second-strand DNA was discontinuous. TdT can add homopolymer tails at internal nicks or gaps (18); tailing at discontinuities in the second-strand product would result in clones deleted for 3′ sequences, but containing information proximal to the 5′ end of the RNA. In contrast, when the cDNA synthesized in the absence of DNA ligase was cloned into pUC9 by the addition of linkers, a dramatic increase in the size of the cloned inserts was observed (Figure 6). Six of 18 inserts were greater than 6 kb long, and 4 of these contained an EcoR I site (Figure 6). In this case, the presence of DNA ligase in the second-strand reaction shifted the size distribution of the cloned cDNA even further towards full-length. Eleven of 19 inserts were greater than 6 kb in length, and 8 of these contained an EcoR I site. Considering that DNA ligase had no effect on the apparent size of second-strand cDNA, it is difficult to explain this shift towards longer cloned inserts in the presence of ligase. It is clear that the method of choice for cloning cDNA derived from long mRNAs by the Gubler and Hoffman method (1) is the addition of linkers. Discontinuities in the second-strand cDNA do not interfere with linker addition and subsequent cloning as long as the ds cDNA has blunt ends.

CONCLUSIONS

Second-strand cDNA synthesis catalyzed by E. coli RNase H and DNA polymerase I (1) results in the loss of sequence information from the 5′ end of an mRNA template; for the templates used here, this corresponded to $\leq$ 21 nucleotides. This loss can be prevented by adding RNase H after, rather than with, DNA polymerase I in the second-strand reaction. Addition of E. coli DNA ligase to RNase H-DNA polymerase I-mediated second-strand cDNA synthesis enhances the recovery of near full-length cloned cDNA from long RNA, particularly when the cDNA is cloned by addition of linkers.

ACKNOWLEDGEMENTS

We thank M. Noon and H. Ley for technical assistance, and M. Kotewicz and D. Rabussay for critical reading of the manuscript.

REFERENCES

1. Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263-269.
2. Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161-170.
3. Aviv, H. and Leder, P. (1972) Proc. Nat. Acad. Sci. USA 69, 1408-1412.
4. Challberg, M.D. and Englund, P.T. (1980) Methods Enzymol. 65, 39-43.
5. Huynh, T.V., Young, R.A., and Davis, R.W. (1985) In Glover, D.M. (ed), DNA Cloning, A Practical Approach, Vol. I, IRL Press, Washington, DC, pp. 49-78.
6. Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.
7. Deng, G. and Wu, R. (1981) Nucl. Acids Res. 9, 4173-4188.
8. Deng, G. and Wu, R. (1983) Methods Enzymol. 100, 96-116.
9. Carmichael, G.G. and McMaster, G.K. (1980) Methods Enzymol. 65, 380-391.
10. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning, A Laboratory Handbook, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
11. Maxam, A.M. and Gilbert, W. (1980) Methods Enzmol. 65, 499-560.
12. Fraenkel-Conrat, H. (1968) Methods Enzymol. 12B, 224-227.
13. Randerath, K. and Randerath, E. (1967) Methods Enzymol. 12A, 323-347.
14. Keller, W. and Crouch, R. (1972) Proc. Nat. Acad. Sci. USA 69, 3360-3364.
15. Leis, J.P., Berkower, I., and Hurwitz, J. (1973) Proc. Nat. Acad. Sci. USA 70, 466-470.
16. Omer, C.A. and Faras, A.J. (1982) Cell 30, 797-805.
17. Resnick, R., Omer, C.A., and Faras, A.J. (1984) J. Virol. 51, 813-821.
18. Williams, J.G. (1981) in Williamson, R. (ed), Genetic Engeering, Vol. 1, Academic Press, New York, p. 29.